UNIVERSIDAD
DE GRANADA

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Facultad de Ciencias

Doble Grado en Ingeniería Informática y Matemáticas

TRABAJO DE FIN DE GRADO

Optimización de redes neuronales

Presentado por:
Blanca Cano Camarero

Tutores:
Juan Julián Merelo Guervós
Arquitectura y tecnología de computadores

Francisco Javier Meri de la Maza
Análisis matemático

Curso académico 2021-2022

Optimización de redes neuronales

Blanca Cano Camarero

Blanca Cano Camarero *Optimización de redes neuronales*.
Trabajo de fin de Grado. Curso académico 2021-2022.
Responsables de tutorización

Juan Julián Merelo Guervós
Arquitectura y tecnología de computadores

Francisco Javier Meri de la Maza
Análisis matemático

Doble Grado en Ingeniería Informática y Matemáticas

Escuela Técnica Superior
de Ingenierías Informática
y de Telecomunicación
Facultad de Ciencias

Universidad de Granada

Resumen

Objetivo inofrmática: Elegir un framework común para trabajar con redes neuronales así como una serie de problemas de complejidad media, tales como spambase. Establecer una línea base examinando los resultados obtenidos con la configuración base a la hora de entrenar este tipo de redes neuronales y el resultado obtenido. A partir de esa línea base, testear las diferentes restricciones, cambios en representación y suposiciones deducidos en la parte matemática para ver qué influencia tienen en la velocidad, en el resultado, o en ambos.

Objetivo matemáticas: El objetivo de esta parte es doble, en primer lugar, se propone analizar con detalle las demostraciones de algunos resultados de aproximación universal de redes neuronales para funciones continuas. En segundo lugar se propone realizar un estudio de la posible optimización de redes neuronales concretas en base a los resultados obtenidos empíricamente en la parte informática. Se tratará de modelizar matemáticamente dichos resultados y de obtener mejoras en la convergencia de las aproximaciones imponiendo, si es necesario, hipótesis más restrictivas en algunos de los elementos de las redes neuronales que se correspondan con su uso en la práctica.

Libros: [1] Abu-Mostafa, Y.S. et al.: Learning From Data. AMLBook, 2012. [2] G. Cybenko, Approximations by superpositions of a sigmoidal function, Math. Contro Signal Systems 2 (1989), 303-314. [3] J. Conway, A Course in Functional Analysis, 2nd Edition, Springer-Verlag, 1990. [4] A. Géron, Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly, 2019. [5] K. Hornik, M Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989), 359-366. [6] W. Rudin, Real and complex analysis. McGraw-Hill Book Co., New York-Toronto, Ont.-London 1966.

PALABRAS CLAVE: redes neuronales LSTM series temporales selección de modelos validación selección de hiperparámetros detección de anomalías detector perturbación

Summary

KEYWORDS: neural networks LSTM time series model selection validation hyper-parameters selection anomaly detection detector perturbation

Índice general

Índice de figuras

Índice de tablas

1. Introducción

Definición 1.1 (Polinomios de Bernstein)**.** Dada cierta función $f : [0, 1] \to \mathbb{R}$, se define el n-ésimo polinomio de Bernstain para f como

$$B_b(x) = B_n(x; f) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

Parte I.

Matemáticas

Esto es un texto nuevo

2. Polinomios de Bernstein

Definición 2.1 (Polinomios de Bernstein)**.** Dada cierta función $f : [0,1] \to \mathbb{R}$, se define el n-ésimo polinomio de Bernstain para f como

$$B_n(x) = B_n(x; f) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\left(\binom{n}{k}\right) x^k (1-x)^{n-k}.$$

La intuición que se esconde tras esta definición es la siguiente: Se pretende aproximar la función f a través de los puntos $\frac{k}{n}$ con $n \in \mathbb{N}$ fijo y $k \in \{0, ..., n\}$. De tal forma que si evaluamos $B_n(x)$ con x lo suficientemente próximo a $\frac{k}{n}$ la imagen de $B_n(x)$ se acerque a $f\left(\frac{k}{n}\right)$.

Para ello recordaremos el teorema del Binomio de Newton:

Teorema 2.1 (Binomio de Newton)**.** *Cualquier potencia de un binomio $x + y$ con $x, y \in R$, puede ser expandido en una suma de la forma*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Así pues en virtud de esta igualdad y puesto que nuestro dominio de definición de f es $[0,1]$, para cualquier $x \in [0,1]$.

Tenemos

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \tag{2.1}$$

Multiplicamos ahora en ambos lados por $f(x)$

$$f(x) = \sum_{k=0}^{n} f(x) \binom{n}{k} x^k (1-x)^{n-k}$$

De manera razonable habríamos encontrado una forma sensata de definir el polinomio buscado, faltaría por ver si efectivamente *aproxima bien*, es decir estudiar la convergencia.

Las diferencia entre $f(x)$ y $B_n(x)$ es de

$$|f(x) - B_n(x)| = \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \tag{2.2}$$

Comp Observando esta ecuación (2.2) ya la intuición nos hace pensar que sea convergente, veremos efectivamente que son uniformemente convergentes.

Teorema 2.2 (Teorema de aproximación de Bernstein)**.** *Sea f una función continua en un intervalo I con imágenes en los reales. La secuencia de polinomio de Bernstein 2.1 converge uniformemente a f en I.*

Recordaremos antes la definición de convergencia uniforme:

Definición 2.2 (Convergencia uniforme para funciones reales)**.** Dado E un conjunto y $\{f_n\}_{n\in\mathbb{N}}$ una sucesión de funciones de E a los reales; se dice que dicha sucesión converge uniformemente si para cualquier $\varepsilon > 0$ existe un número natural m tal que para todo $x \in E$ y cualquier natural n que cumpla $n \geq m$ se tiene que

$$|f_n(x) - f(x)| < \varepsilon$$

Comencemos pues con la demostración del teorema 2.2.

Demostración. Sin pérdida de generalidad supondremos que $I = [0,1]$, como veremos esto no es restrictivo ya que si I fuera un intervalo cerrado existiría un homeomorfismo H tal que $H^*(I) = [0,1]$ y podríamos trabajar con $H \circ f$ la cual respetaría todas los argumentos utilizados en la demostración.

Si I fuera un abierto consideraríamos su cierre y aplicaríamos el razonamiento anterior. De esta manera los supremos e ínfimos se mantendrían, ahora como mínimos y máximos y no se alteraría de ninguna manera la continuidad.

Tras la aclaración anterior podemos comenzar.

Sea $\varepsilon > 0$ queremos probar que existirá un $m_\varepsilon \in \mathbb{N}$ tal que para cualquier $x \in I$ e $n \geq m_\varepsilon$ se tenga que $|f(x) - B_n(x)| < \varepsilon$.

Para ello por estar f definida en un intervalo, se tienen dos consecuencias claves:

1. Está acotada, supongamos por $M \in \mathbb{R}$, esto es $|f(x)| \leq M$.

2. En virtud del teorema de Heine-Cantor f es uniformemente continua, es decir: para cualquier $\varepsilon > 0$ existirá un δ_ε tal que para cualesquiera $x, y \in I$ que cumplan $|x - y| < \delta_\varepsilon$ entonces $|f(x) - f(y)| < \varepsilon$

Dado $N \in \mathbb{N}$ fijo pero arbitrario, para $\{k|k \in \{1, ..., N\}\ \frac{k}{N} \in I$ y tomando $x \in I$ podemos acotar

$$\left| f(x) - f\left(\frac{k}{N}\right) \right| \leq |f(x)| + \left| f\left(\frac{k}{N}\right) \right| \leq 2M$$

Fijado un x del dominio, se tienen las siguientes particiones de índices:

$$\mathcal{A}_{x,N} = \{k|k \in \{1, ..., N\} \text{ y } |x - \frac{k}{N}| < \delta_\varepsilon\}$$

Donde el δ_ε se ha obtenido de la observación 2 tomando como ε el buscado para la convergencia uniforme.

Sea $\varepsilon > 0$ y tomemos δ_ε el delta de la definición de uniformemente continuo.

Por otro lado se define

$$\mathcal{B}_{x,N} = \{1, ..., N\} - \mathcal{A}_{x,N}$$

Podemos elegir m de manera conveniente de tal forma que

$$n \geq sup\left\{(\delta_\varepsilon)^{-4}, \frac{M^2}{\varepsilon^2}\right\}$$

y agruparemos los términos de la sumatoria en dos partes, para los que $|x - \frac{k}{n}| < n^{-\frac{1}{4}} \leq \delta_\varepsilon$.

Para éstos, se tiene por Newton

$$\sum_{k=1}^{n} \varepsilon \binom{n}{k} x^k (1-x)^{n-k} \leq \varepsilon \sum_{k=1}^{n} \binom{n}{k} x^k (1-x)^{n-k} = \varepsilon$$

Para el resto de términos, aquellos cuya k cumpla que $|x - \frac{x}{n}| \geq n^{\frac{-1}{4}}$, se tiene que $|x - \frac{x}{m}| \geq n^{\frac{-1}{2}}$.

Queda el resto de subíndices resulta acotada por, deberemos de ver que para una determinada N, esa desigualdad se queda en épsilon

$$\sum_{k \in \mathcal{B}_{x,N}} 2M \binom{n}{k} x^k (1-x)^{n-k}$$

Para acotar esto nos va a costar trabajar un poco más

$$\begin{aligned}
&\sum_{k \in \mathcal{B}_{x,N}} 2M \binom{N}{k} x^k (1-x)^{N-k} \\
&= 2M \sum_{k \in \mathcal{B}_{x,N}} \frac{(x - \frac{k}{N})^2}{(x - \frac{k}{N})^2} \binom{N}{k} x^k (1-x)^{N-k} \\
&\leq 2M\sqrt{N} \sum_{k \in \mathcal{B}_{x,N}} (x - \frac{k}{N})^2 \binom{n}{k} x^k (1-x)^{N-k}
\end{aligned}$$

Tengamos ahora presenten las siguientes igualdades

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} = \frac{k}{n}\binom{n}{k} \tag{2.3}$$

$$\binom{n-2}{k-2} = \frac{(n-2)!}{(k-2)!(n-2-(k-2))!} = \frac{k(k-1)}{n(n-1)}\binom{n}{k} \tag{2.4}$$

Partiendo de la igualdad (2.1):

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}$$

Reemplazamos la n por $n-1$ y la k por j y tenemos

$$1 = \sum_{j=0}^{n-1} \binom{n-1}{j} x^j (1-x)^{(n-1)-j}$$

Multiplicamos por x y aplicamos la igualdad (2.3) resultando

$$x = \sum_{j=0}^{n-1} \frac{j+1}{n} \binom{n}{j+1} x^{j+1} (1-x)^{(n-(j+1)}$$

Renombramos $k = j+1$, por lo que resulta

$$x = \sum_{k=1}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k}$$

Como el término con $k = 0$ es nulo podemos añadirlo a la sumatoria

$$x = \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k} \tag{2.5}$$

Haremos ahora un razonamiento similar sustituyendo n por $n-2$

Partiendo de (2.1) se tiene que

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}$$

Reemplazamos la n por $n-2$ y la k por j y tenemos

$$1 = \sum_{j=0}^{n-2} \binom{n-2}{j} x^j (1-x)^{(n-2)-j}$$

Multiplicamos por x^2 y aplicamos la igualdad (2.4) resultando

$$x^2 = \sum_{j=0}^{n-2} \frac{(j+2)(j+1)}{n(n-1)} \binom{n}{j+2} x^{j+2} (1-x)^{(n-(j+2))}$$

Renombramos $k = j+2$, por lo que resulta

$$x^2 = \sum_{k=2}^{n} \frac{k(k-1)}{n(n-1)} \binom{n}{k} x^k (1-x)^{n-k}$$

Como con los términos $k=0$ y $k=1$ se anula, podemos añadir dichos índices sin modificar la suma

$$x^2 = \sum_{k=0}^{n} \frac{k(k-1)}{n(n-1)} \binom{n}{k} x^k (1-x)^{n-k}$$

Podemos reescribir la ecuación resultando:

$$(n^2-n)x^2 = \sum_{k=0}^{n} (k^2-k) \binom{n}{k} x^k (1-x)^{n-k} \tag{2.6}$$

Recordemos que nuestro objetivo era acotar 2

Para ello vamos a sumar las dos expresiones que hemos obtenido (2.5) y (2.6) resultando

$$(n^2-n)x^2 + nx = \sum_{k=0}^{n} ((k^2-k)+k) \binom{n}{k} x^k (1-x)^{n-k}$$

Dividimos todo entre n.

$$(1-\frac{1}{n})x^2 + \frac{1}{n}x = \sum_{k=0}^{n} \left(\frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k} \tag{2.7}$$

A continuación sumamos a la igualdad (2.7) la ecuación (2.1) multiplicada por x^2 y la ecuación (2.5) multiplicada por $-2x$ resultando:

$$(1-\frac{1}{n}+1-2)x^2 + \frac{1}{n}x = \sum_{k=0}^{n} \left(\left(\frac{k}{n}\right)^2 + x^2 - 2x\right) \binom{n}{k} x^k (1-x)^{n-k}$$

Factorizando en ambos miembros resulta

$$\frac{1}{n}x(1-x) = \sum_{k=0}^{n}\left(x-\frac{k}{n}\right)^2\binom{n}{k}x^k(1-x)^{n-k} \tag{2.8}$$

Gracias a (2.8) acabos de encontrar una cota para (2)

$$2M\sqrt{N}\sum_{k\in\mathcal{B}_{x,N}}(x-\frac{k}{N})^2\binom{N}{k}x^k(1-x)^{N-k} = 2M\sqrt{N}\frac{1}{N}x(1-x) = 2M\frac{1}{\sqrt{N}}x(1-x)$$

Además como $x(1-x)$ alcanza un máximo absoluto en $x=\frac{1}{2}$ luego concluimos que

$$|f(x)-B_N(x)| = \sum_{k=0}^{N}\left|f(x)-f\left(\frac{k}{N}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} \leq M\frac{1}{2\sqrt{N}}$$

Con $M = max\{|f(x)| : x \in I\}$.
Por que tomando como $m_\epsilon = \lceil M\frac{1}{2\sqrt{\epsilon}}\rceil$ habremos probado la convergencia uniforme buscada.

□

3. Teorema de Stone-Weierstass

Teorema 3.1 (Teorema de Stone-Weierstass). *Sea K un subconjunto compacto de $\mathbb{R}^p$ y sea $\mathcal{A}$ una colección de funciones continuas de K a $\mathbb{R}$ con las siguientes propiedades:*

1. *La función constantemente uno, definida como $e(x) = 1$, para cualquier $x \in K$ pertenece a $\mathcal{A}$.*
2. *Cerrado para sumas y producto para escalares. Si f, g pertenece a $\mathcal{A}$, entonces $\alpha f + \beta g$ pertenece a $\mathcal{A}$.*
3. *Cerrado para producto, fg pertenece a $\mathcal{A}$.*
4. *Separación de K, es decir si $x \neq y$ pertenecientes a K, entonces existe una función f en $\mathcal{A}$ de tal manera que $f(x) \neq f(y)$.*

Se tiene que toda función continua de K a $\mathbb{R}$ puede ser aproximada en K por funciones de $\dashv$.

Demostración. $\square$

Bibliografía

Las referencias se listan por orden alfabético. Aquellas referencias con más de un autor están ordenadas de acuerdo con el primer autor.

Agradecimientos

Agradezco a